Simplified Dependency Annotations with GFL-Web

Michael T. Mordowanec, Nathan Schneider, Chris Dyer, Noah A. Smith

Treebanking is expensive. What if it weren’t?

This is the motivating question behind GFL-Web, which uses the Fragmentary Unlabeled Dependency Grammar, or FUDG (Schneider et al. 2013), to make syntactic representations easier on annotators. FUDG is written using the Graph Fragment Language or GFL, a simple ASCII-based notation, which annotators can master in a few minutes.

This framework allows annotators to work quickly, underspecify where necessary, and still create a useful treebank. Annotators can leave out sections of the text, group words whose precise relationship is uncertain or time-consuming, and receive instant feedback in the form of a graph visualization.

Notation:

| The > dog | dog < The                                    | Unlabeled dependency edge |
| [President Barack Obama] | Groups multiword units                       |
| {a silver} > dollar | All elements inside become children of same parent node. |
| (even though) | Underspecified relationship                  |
| The > dog** < ran | Marks a root node.                          |

Input:

Texas Rangers are in the World Series! Go Rangers !!!!!!!! http://fb.me/D2LsXBJx

Annotation:

[Texas Rangers~1] > are** < in
in < (the > [World Series])
Go** < Rangers~2